Vershinina Ekaterina Victorovna

Student

Ural Federal University

Russia, Ekaterinburg

Academic supervisor: Kovaleva Alexandra Georgievna

VIDEO IMAGE ANALYSIS METHODS THAT USE CONVOLUTIONAL NEURAL NETWORKS

Abstract. *The paper is devoted to the introduction of the basic principles of working with convolutional neural networks, especially in the tasks related to the identification of moving objects. The study also provides information about the operation of convolutional layers and their interaction with each other. At the moment, the analysis of a large number of video images has to be carried out manually, which takes a long time and is impossible without human intervention. Neural networks allow independent analysis and identification on the basis of algorithms embedded in the program.*

Keywords: *video analysis, convolutional neural networks, convolutional layers, identification, moving objects.*

Вершинина Екатерина Викторовна

Студент

Уральский федеральный университет

Россия, г. Екатеринбург

Научный руководитель: Ковалева Александра Георгиевна

ИССЛЕДОВАНИЕ МЕТОДОВ АНАЛИЗА ВИДЕОИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

Аннотация. *Статья посвящена ознакомлению с основными принципами работы со сверточными нейронными сетями, в особенности в задачах связанных с идентификацией движущихся объектов. Также в работе представлена информация о работе сверточных слоев и их взаимодействии друг с другом. На данный момент, анализ большого количества видеоизображений приходится проводить вручную, что занимает большое время. Нейронные сети позволяют проводить анализ и идентификацию самостоятельно на основе алгоритмов заложенных в программу.*

Ключевые слова: *видеоанализ, сверточные нейронные сети, сверточные слои, идентификация, движущиеся объекты.*

Convolutional neural networks (CNN) are an architecture aimed at efficient pattern recognition, which is part of the deep learning technology. The name «convolutional neural network» indicates that this network uses a mathematical operation called convolution. Convolution is a specialized type of linear operation. Thus, convolutional networks are neural networks that use convolution instead of matrix multiplication in at least one of their layers. CNN have been actively used in the last decade for tasks related to the recognition of various images. This allows developers to implement this technology in various areas of life, especially if they are associated with probabilistic identification and have a large number of different parameters.

One of the applications of convolutional neural networks is the identification of athletes in broadcast sports videos. Multi-camera sports video analysis allows getting data about the match and the positions of players on the field during it. The resulting analysis allows implementing various applications, including improving the broadcast of sports videos, reconstructing a 3D match for a more complete view of all the events taking place on it, and providing interactive content for the audience. Also, the data obtained as a result of CNN work simplifies the collection of game data to support coaches in conducting tactical analysis, and allows you to resolve controversial game moments with less probability of error.

At the moment, there is a large number of studies related to the identification of objects in space. Some studies are related to the creation of models that are used for identification [4], some are related to the creation of clearer images of moving objects [5], and machine learning is also the object of research [2]. This increases the amount of data received during the match, as well as greatly simplifies their analysis. Thus, the purpose of the study is to analyse researches related to the use of convolutional neural networks for video image processing. The data obtained as a result of the analysis allow us to form a general solution.

We present basic information about convolutional neural networks, the nature of their application, and the main algorithms. This guarantees better understanding of the theoretical component of the issue, understanding how the processes of convolution and analysis of the collapsed image are implemented.

Modern deep learning is the basis of the services of many companies: Facebook uses neural network algorithms for automatic affixing of tags Google — to search among the photos of the user, Amazon — to generate product recommendations, and Instagram for search infrastructure.

The idea of CNN is to alternate a large number of convolutional layers. Each layer solves a specific problem, and the result of the algorithm is a demonstration of a class or a group of probability classes that best characterize this image. The filter performs convolution, which is moving through the water image, multiplies the filter values by the original pixel values of the image. All these multiplications are added together. And the result is a single number. Then the process is repeated in each position (Figure 1). The next step is to move the filter to the right by one, then another one to the right, and so on.

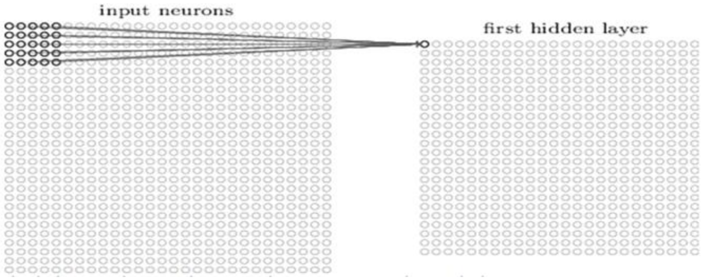

Figure 1. – Visualization of filter convolving around an input volume and producing an activation map [1].

Each filter can be considered as a property identifier. The ID property can be curves, borders, and simple colors (Figure 2).

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

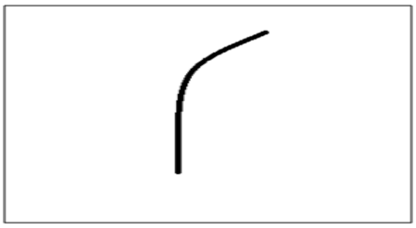

Figure 2. –Pixel representation of filter and visualization of a curve detector filter.

When the filter is located in the upper-left corner of the input image, it multiplies the filter values by the pixel values of this area.

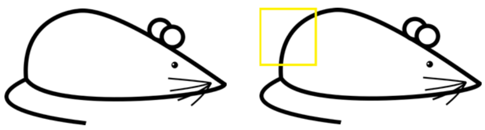

Figure 3- Original image and visualization of the filter on the image [3].

If the input image has a shape that is roughly similar to the curve that this filter represents, and all the multiplied values are added together, the result is a large number (Figure 4). This high value indicates that it is possible that something similar to the curve is present in the image, and this probability activated the filter. The filter will now be moved to a different location (Figure 5).

Visualization of the receptive field

0	0	0	0	0	0	30
0	0	0	0	50	50	50
0	0	0	20	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0

Pixel representation of the receptive field

*

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

Pixel representation of filter

Multiplication and Summation = (50*30)+(50*30)+(50*30)+(20*30)+(50*30) = 6600 (A large number!)

Figure 4. – The result of the filter in the first position [3].

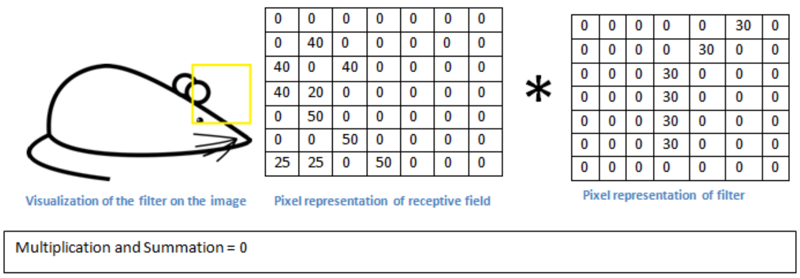

Figure 5. – The result of the filter in the second position [3].

The resulting value is much smaller than the previous one. This is due to the fact that the curve that the filter is trying to find is not similar to the curve located in this sector of the image.

Algorithms based on convolutional neural networks work well with tasks related to the identification of moving objects. The result of an algorithm that improves the quality of images obtained from shooting a large number of moving objects is discussed below. This algorithm was presented by a group of scientists from Chinese laboratories [5]. The figures demonstrate three States for comparison. The first is data point targets generated by Range Doppler Algorithm (RD) (Range-Doppler algorithm is one of the most popular synthetic aperture radar processing algorithms). The second is the image obtained as a result of the algorithm. And the third is the ideal result that can be obtained. Data is provided for moving objects at different speeds. Figures 6 and 7 demonstrate the results for stationary objects. Figures 8 and 9 show the results for objects moving at a speed of 12 m/s.

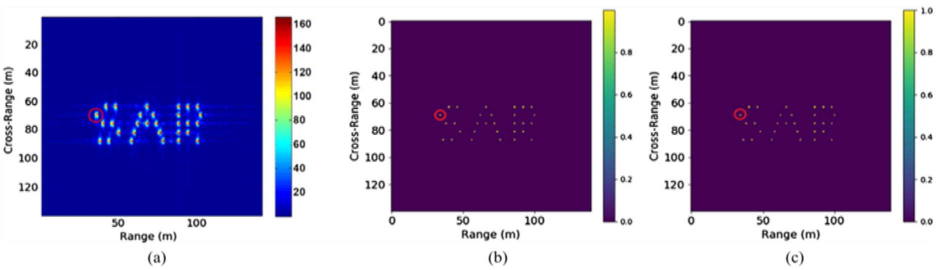

Figure 6. - Imaging results of the stationary point target. (a) RD algorithm. (b) U-net output (c) Ideal scattering point model [5].

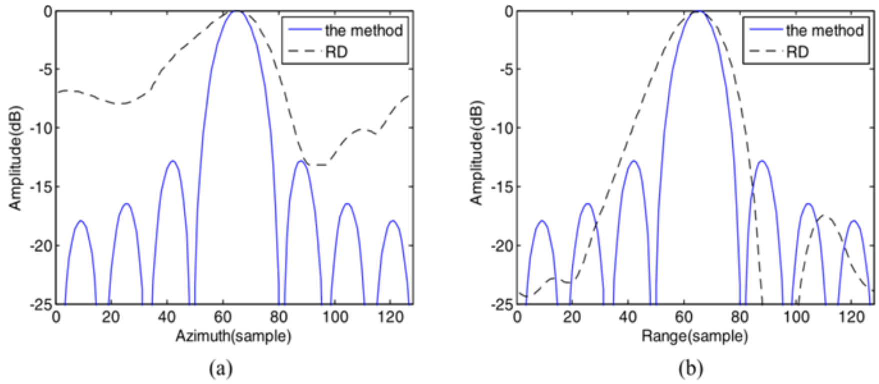

Figure 7- Profile comparison of the scattering point marked by the red circle in Fig. 6. (a) Comparison in azimuth. (b) Comparison in range [5].

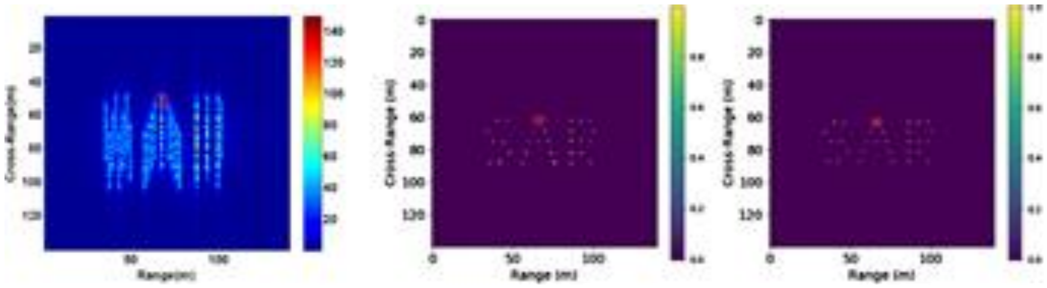

Figure 8. - Imaging results of the moving point target with azimuth velocity of 12 m/s. (a) RD algorithm. (b) U-net output. (c) Ideal scattering point model [5].

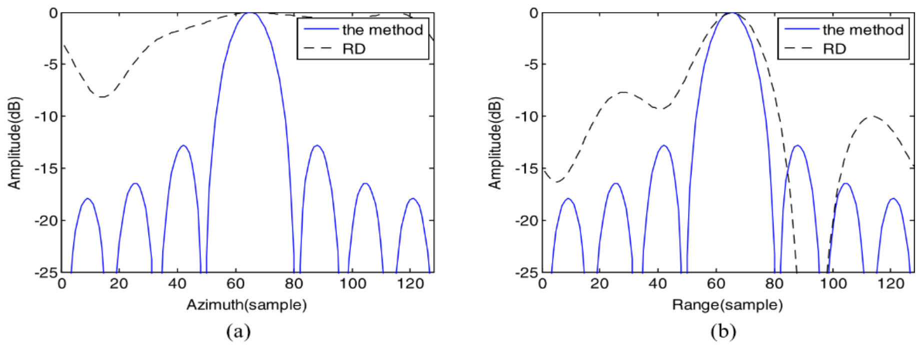

Figure 9. - Profile comparison of the moving point marked by the red circle in Figure 8. (a) Comparison in azimuth. (b) Comparison in range [5].

The resulting images make clear that the algorithm based on convolutional neural networks allows getting fairly accurate images of moving objects, neutralizing extraneous noise. At the same time, the result is close to the ideal indicators, both for stationary objects and for objects that move at high speed.

Convolutional neural networks are defined in the paper. The general logic of convolution matrices is analyzed. We demonstrate how convolutional layers work and how to analyse the information received. Using the example of an algorithm that receives an image of moving objects, we demonstrate the implementation of the algorithm based on convolutional neural networks. The knowledge gained in the study allows us to continue deeper study of this topic and contribute to further research.

REFERENCES

1. Adit Deshpande. A Beginner's Guide To Understanding Convolutional Neural Networks - 2020 - Text: electronic URL: https://adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks/ -(Reference date 26.12.2020).

2. Ivet Rafegas, Maria Vanrell, Luis A. Alexandre, Guillem Arias. Understanding Trained CNNs by Indexing Neuron Selectivity// Pattern Recognition Letters/ Text: electronic. – 2019- URL: https://www.sciencedirect.com/science/article/abs/pii/S0167865519302909 -(Reference date 26.12.2020).

3. Michael Nielsen. Neural Networks и Deep Learning Text: electronic. – 2019 – URL: http://neuralnetworksanddeeplearning.com/chap1.html -(Reference date 26.12.2020).

4. RuihengZhang, LingxiangWu, YukunYang, WannengWu, YueqiangChen, MinXu. Multi-camera multi-player tracking with deep player identification in sports video //Pattern Recognition 2020 - Text: electronic. – URL: https://www.sciencedirect.com/science/article/abs/pii/S0031320320300650 -(Reference date 26.12.2020).

5. Zhi-junLu, QiQin, Hong-yinShi, HaoHuang - SAR moving target imaging based on convolutional neural network. //Digital Signal Processing- Text: electronic. – URL: https://www.sciencedirect.com/science/article/abs/pii/S1051200420301779 -(Reference date 26.12.2020).